In fact, halogenation and nitration of the ring is observed to occur most readily at the *ortho* and *para* positions with respect to the C=O group, i.e. at C_2 and C_4. Disubstitution in these positions is possible but it is very difficult to go beyond this stage. This structure analysis may therefore be said to provide direct evidence in one particular case of the existence of an effect which has been postulated to explain the reactivity of certain substituted aromatic compounds. Furthermore, the evidence is that this effect exists separately from the chemical reaction itself, and is not entirely brought about by the approach of reacting groups.

I should like to thank Prof. Sir Lawrence Bragg and Dr W. H. Taylor for facilities provided for this work and for their interest in it. I am greatly indebted to Dr W. Cochran for his encouragement and for much helpful advice. I also wish to thank Mrs L. Remnant for assistance with computation. I acknowledge gratefully the award of post-graduate scholarships by the University of New Zealand and the British Council during the tenure of which this work was carried out, and thank the Council of Canterbury University College, New Zealand, for the granting of three years leave of absence.

References

ALBRECHT, G. (1939). *Rev. Sci. Instrum.* **10**, 221.
ALLEN, P. W. & SUTTON, L. E. (1950). *Acta Cryst.* **3**, 57.
BACON, G. E. (1952). *Acta Cryst.* **5**, 492.
COCHRAN, W. (1948). *Acta Cryst.* **1**, 54.
COCHRAN, W. (1950). *Acta Cryst.* **3**, 268.
COCHRAN, W. (1951). *Acta Cryst.* **4**, 81.
COCHRAN, W. (1953). *Acta Cryst.* **6**, 260.
COX, E. G. & CRUICKSHANK, D. W. J. (1948). *Acta Cryst.* **1**, 92.
ELDERFIELD, R. C. (1950). *Heterocyclic Compounds*. New York: Wiley.
GOLDSCHMIDT, G. H. & LLEWELLYN, F. J. (1950). *Acta Cryst.* **3**, 294.
HUGHES, E. W. (1941). *J. Amer. Chem. Soc.* **63**, 1737.
JAMES, R. W. & BRINDLEY, G. W. (1931). *Phil. Mag.* (7), **12**, 81.
MCDONALD, T. R. R. & BEEVERS, C. A. (1952). *Acta Cryst.* **5**, 654.
MCWEENY, R. (1951). *Acta Cryst.* **4**, 513.
PATTERSON, A. L. (1949). *Acta Cryst.* **2**, 339.
PAULING, L. (1940). *The Nature of the Chemical Bond*, 2nd ed. Ithaca: Cornell University Press.
REMICK, A. E. (1943). *Electronic Interpretations of Organic Chemistry*. New York: Wiley.
VAUGHAN, P. & DONOHUE, J. (1952). *Acta Cryst.* **5**, 530.

Acta Cryst. (1953). **6**, 600

The Strength of the 10 Å Reflexions in Haemoglobin

BY F. H. C. CRICK

The Medical Research Council Unit for the Study of the Molecular Structure of Biological Systems, The Cavendish Laboratory, Cambridge, England

(*Received* 14 *February* 1953)

A method is described for calculating the total intensity of X-ray reflexions at spacings near 10 Å that one would expect from models of a protein consisting mainly of α-helices, not necessarily parallel. A comparison with Perutz's observed data for haemoglobin shows that the observed intensity is compatible with such models. The comparison can be made only with moderate accuracy.

Introduction

All recent work (see Crick (1952*b*) for references) has confirmed that the structure of the synthetic polypeptides is based on the α-helix of Pauling, Corey & Branson (1951). Very recently, Crick (1952*b*) has pointed out that packing considerations might deform an α-helix so that it became a coiled-coil, and simultaneously Pauling & Corey (1953) have produced a definite model of this type for α-keratin (though based on different assumptions). This model explains qualitatively the main features of the α-keratin pattern, namely the 1·5 Å and 5·15 Å meridian reflexions, and the reflexions of about 10 Å on and near the equator. It is thus probable that the fibrous proteins of the α type are largely based on the α-helix.

The structure of the globular proteins is less satisfactory. Pauling & Corey (1951*b*) originally suggested that the radial distribution function of haemoglobin, which they derived in an approximate manner from Perutz's published data (Perutz, 1949), appeared to resemble that expected from the α-helix. This approach has been taken considerably further by Riley & Arndt (1952), who have compared the observed radial distribution function for bovine serum albumin in the form of a dry powder, with Pauling & Corey's

theoretical curve for the α-helix. The result is suggestive, but not conclusive, and further work along these lines, both experimental and theoretical, is awaited with interest.

Perutz (1951) has observed a 1·5 Å reflexion from haemoglobin, but it was weaker than that observed for the synthetic polypeptide. For other globular proteins there is no published information available.

Perutz's early work (Perutz, 1949) on haemoglobin suggested that the structure might consist largely of rods of high electron density parallel to the a axis of the crystal. However, later more quantitative work by Bragg, Howells & Perutz (1952) showed that this was unlikely as the projection along the a axis gave peaks smaller than one would have expected. At about the same time Crick (1952a) showed from a quantitative examination of Perutz's three-dimensional Patterson that it was quite incompatible with any model consisting mainly of long straight α-helices parallel to the a axis. The fact that three-dimensional data were used made the argument more conclusive than that of Bragg, Howells & Perutz, which was based on a two-dimensional projection only.

It is difficult to attack a crystallographic problem of the complexity of protein structure without some simplifying idea. The great structural plausibility of the α-helix, its success in predicting the general nature of the X-ray diffraction pattern of all known synthetic polypeptides in the α-form, and the recent plausible explanation of α-keratin in terms of curved α-helices, all strongly suggest that it is the key to protein structure. It cannot therefore be lightly abandoned. Since haemoglobin cannot consist mainly of straight parallel α-helices, the next most plausible model to consider is one consisting of α-helices packed in a non-parallel manner. There are at least two reasons why this might happen. First, although haemoglobin does not consist chemically of four myoglobins, it may well consist of four 'myoglobin' units from the structural points of view. If these sub-units were not all parallel this might produce a dispersion of the axial directions of the α-helices.

Secondly, since α-helices of the same sense may well pack better when at an angle of 20° away from parallel (Crick, 1952b) the α-helices within the sub-units (if they exist) may not be strictly parallel to each other.

Before attacking the problem of the haemoglobin structure in detail, therefore, one would like to know if one can glean anything from the general distribution of intensities, so that time is not spent in pursuing quite impossible models.

The radial intensity distribution

It is the 10 Å reflexions which produce the appearance of rods seen end-on in the a projection of haemoglobin (Bragg, Howells & Perutz, 1952). An inspection of the complete diffraction pattern shows that there are reflexions of about this spacing and of comparable strength in other directions, and suggests that one should consider the *average* strength of all the 10 Å reflexions.

The curve of average intensity plotted against the reciprocal spacing is shown in Fig. 1. This has been obtained from Perutz's three-dimensional intensities, and is the same as his Fig. 24 (Perutz, 1949) but recalculated at closer intervals. It has been put on an absolute scale using data of Perutz.

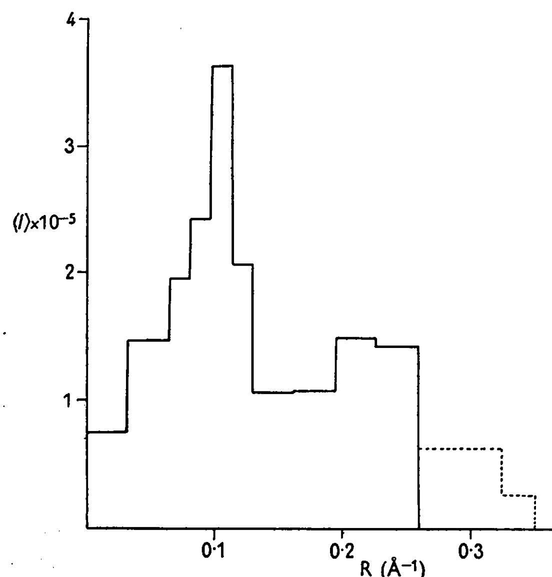

Fig. 1. The observed distribution of intensities for haemoglobin, from data of Perutz (broken line—estimated only).

It is at once apparent that the distribution of intensities does not correspond to a 'random' distribution of atoms, and in fact the whole idea of a 'random' distribution is useless for these large molecules. The peakiness of Fig. 1 clearly indicates that there is substructure of some sort within the haemoglobin molecule.

The simplest way to tackle the problem might appear to be a quantitative examination of the radial distribution curve obtainable from Fig. 1. This is the method of Riley & Arndt. It cannot easily be used here, since the recorded intensities do not extend far enough. This produces diffraction effects and also makes it impossible to obtain an accurate estimate of the temperature factor. Thus it is difficult to use the reflexions in the 4–5 Å region without getting out of the calculation what one put in. It is therefore necessary to concentrate on the 10 Å region, and the question one wishes to answer is 'Is there enough intensity in the 10 Å region to be compatible with models made up mainly of α-helices?'.

It is clear that one cannot merely consider the peak height of Fig. 1, since this will partly depend on how the α-helices are arranged. What one requires is a weighted sum of the 10 Å region which one can easily

relate to a known aspect of the physical model. Fortunately it is possible to obtain this.

The mathematical method

If ϱ is the electron density, and V the volume of the unit cell, and assuming orthogonal axes for simplicity,

$$\varrho = \frac{1}{V} \sum_{hkl} F_{hkl} \exp\left[2\pi i(hx/a+ky/b+lz/c)\right],$$

whence

$$\frac{\partial \varrho}{\partial x} = \frac{1}{V} \sum_{hkl} (2\pi i h/a) F_{hkl} \left[\exp 2\pi i(hx/a+ky/b+lz/c)\right],$$

and therefore, by a well known theorem,

$$\int \left(\frac{\partial \varrho}{\partial x}\right)^2 dV = \frac{1}{V} \sum_{hkl} 4\pi^2 (h^2/a^2) |F_{hkl}|^2 .$$

Repeating for y and z, we obtain

$$\int \left[\left(\frac{\partial \varrho}{\partial x}\right)^2 + \left(\frac{\partial \varrho}{\partial y}\right)^2 + \left(\frac{\partial \varrho}{dz}\right)^2\right] dV = \frac{4\pi^2}{V} \sum_{hkl} R^2 |F_{hkl}|^2 ,$$

where

$$R^2 = h^2/a^2+k^2/b^2+l^2/c^2 \qquad \text{as usual;}$$

i.e.

$$\int |\text{grad}\, \varrho|^2 dV = \frac{4\pi^2}{V} \sum_{hkl} R^2 |F_{hkl}|^2 . \qquad (1)$$

In words, we can obtain the average value of the square of the slope of the electron density by taking the sum of all the intensities weighted by the square of the reciprocal spacing. To make the sum converge we may have to add an artificial temperature factor as a smoothing factor. If we choose this factor so that it smooths out all the details of the atoms and leaves us with only the broad picture of the rods of electron density due to the backbone of the α-helices, we shall get a measure of the amount of up-and-down variation in density at this low resolution by carrying out this summation. Notice three points:

(*a*) F^2_{000} has weight zero, so that one is not swamped by a large unhelpful term in this method.

(*b*) One requires only the intensities; it is not necessary to know the phases of the reflexions.

(*c*) Nevertheless the effects are additive. That is, if one has a structure consisting of two α-helices, one can perform the integral over the electron density of each separately and then add the results together, provided that their (smoothed) electron densities do not overlap. Their relative distance and orientation makes no difference to the integral.

The smoothing function must be chosen so that the sum on the right of equation (1) converges. It must be such that a substantial portion of the sum comes from the 10 Å region, but it must not spread the electron density so much that overlapping between adjacent helices becomes important.

If one smooths by multiplying the intensities by

$$\exp(-\gamma R^2)$$

one is in effect using

$$R^2 \exp(-\gamma R^2)$$

as a weighting function. The value of γ chosen, which was a compromise, was $\gamma = 81$ Å^2. The form of $\gamma R^2 \exp(-\gamma R^2)$ is shown in Fig. 2. It has a maximum at a spacing of 9 Å.

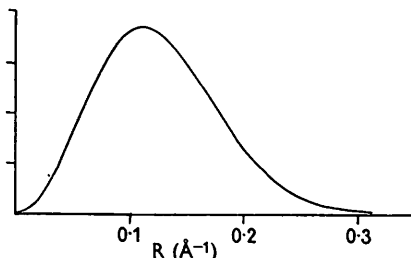

Fig. 2. Plot of $\gamma R^2 \exp(-\gamma R^2)$ for $\gamma = 81$ Å^2.

The method therefore consists of taking some convenient theoretical model for the α-helix, such as one in which all the chains are parallel, and obtaining the right-hand side of equation (1) from the *calculated* intensities. The same sum is carried out using the observed intensities from the protein crystal, and the two figures are then compared.

The theoretical model

This basic model consists of an infinite array of infinite parallel α-helices arranged in a hexagonal lattice, $a = 10{\cdot}4$ Å. The β carbon atom is put in the $C\beta_2$ position (Pauling & Corey, 1951*a*). The remaining side-chains are put in uniformly outside a circle radius 4 Å with sufficient electron density to make the overall average density 0·43 e.Å^{-3}. The calculation was carried out for point atoms, smoothed so that

$$\langle I \rangle = \langle I_0 \rangle \exp(-25R^2) . \qquad (2)$$

From this model the values of $|F_{hkl}|^2$ were calculated. The summation carried out was

$$\frac{4\pi^2}{V} \sum_{hkl} R^2 |F_{hkl}|^2 \exp(-81R^2) . \qquad (3)$$

The result was

$$0{\cdot}8 \text{ e.}^2\text{Å}^{-5} \text{ per residue of } \alpha\text{-helix.}$$

The experimental data

The sum given in (3) was evaluated for haemoglobin from Fig. 1. The result, for the end-centred cell was

$$840 \text{ e.}^2\text{Å}^{-5}.$$

As this cell contains two molecules, the result can be expressed as

$$420 \text{ e.}^2\text{Å}^{-5} \text{ per molecule.}$$

If we divide this by the theoretical value which we have obtained per residue, namely 0·8, we get 525 as the 'effective' number of residues per molecule. The actual number in haemoglobin is about 580 (Tristram, 1949). In other words, the calculated and observed values agree to within 10%.

The agreement is good and in fact it is probably accidental since the method cannot be expected to be so accurate.

Possible errors

The effect of salt can certainly be neglected since it affects only reflexions which make a very small contribution to (3). If we had chosen position $C\beta_1$ instead of $C\beta_2$ this could have made only about 5% difference. A value of 17·5 instead of 25 in equation (2) increases the value of (2) by about 15%. The effects of overlapping are difficult to estimate, since it is not easy to choose an alternative model. An estimate using data from polymethylglutamate, which has $a = 12{\cdot}0$ Å instead of 10·4 Å, suggests that the sum (3) has *decreased* about 10%, but part of this may be due to other differences between the two models.

A detailed examination of the data suggests that the theoretical model does not give enough intensity in the 5 Å region, owing to the side-chains having been slurred unduly. This might produce errors of, say, 20%, but without some idea of how side-chains pack it is difficult to improve the model.

No allowance has been made for the fact that the actual α-helices (if they exist) must be of finite length, and therefore, that part of the polypeptide chain must be used to join adjacent α-helices. It is not even easy to be certain whether this effect will increase or decrease the observed sum.

Discussion

Having made all these allowances for the comparative crudeness of the method, the results of the calculation are nevertheless satisfactory in that they show that it is not impossible for haemoglobin to consist mainly of α-helices. Any argument designed to disprove this from the 10 Å reflexions will have to be more precise and sophisticated than the simple method used here.

On the other hand, the calculation should not be taken to *prove* that the haemoglobin structure is largely α-helices. It merely shows that the 10 Å intensities are not incompatible with such a view, and encourages one to attempt to solve the structure along these lines.

Our method, as explained above, consists in taking a weighted sum of the 10 Å region. It can be seen from Fig. 2 that the weighting function used is fairly broad, and covers a wide range of spacings. The fact that the peak in Fig. 1 is sharp and occurs at about 9 Å spacing gives additional information which has not been made use of here. It suggests that the α-helices in haemoglobin (if they exist) are usually about 10 Å apart laterally, rather than, say, 15 Å.

It would be possible to design a weighting function that was sharper than Fig. 2, but to use it we should have to be more precise about our model. The very wide range of theoretical models which could give the correct value to our summation (3) is both the strength of the method—since we do not have to make our model precise—and its weakness—since the method cannot be used to distinguish between these models.

I should like to thank Dr M. F. Perutz for allowing me to use his three-dimensional haemoglobin data.

References

Bragg, W. L., Howells, E. R. & Perutz, M. F. (1952). *Acta Cryst.* **5**, 136.

Crick, F. H. C. (1952*a*). *Acta Cryst.* **5**, 38.

Crick, F. H. C. (1952*b*). *Nature, Lond.* **170**, 882.

Pauling, L. & Corey, R. B. (1951*a*). *Proc. Nat. Acad. Sci., Wash.* **37**, 235.

Pauling, L. & Corey, R. B. (1951*b*). *Proc. Nat. Acad. Sci., Wash.* **37**, 282.

Pauling, L. & Corey, R. B. (1953). *Nature, Lond.* **171**, 59.

Pauling, L., Corey, R. B. & Branson, H. R. (1951). *Proc. Nat. Acad. Sci., Wash.* **37**, 205.

Perutz, M. F. (1949). *Proc. Roy. Soc.* A, **195**, 474.

Perutz, M. F. (1951). *Nature, Lond.* **167**, 1053.

Riley, D. P. & Arndt, U. W. (1952). *Nature, Lond.* **169**, 138.

Tristram, G. R. (1949). *Advances in Protein Chemistry*, vol. 5. New York: Academic Press.